[EJ] Oral | M (Multidisciplinary and Interdisciplinary) | M-GI General Geosciences, Information Geosciences & Simulations

[M-GI29] [EJ] Data-driven analysis, modeling and prediction in geosciences

convener:Tatsu Kuwatani(Japan Agency for Marine-Earth Science and Technology), Dmitri Kondrashov(University of California, Los Angeles), Hiromichi Nagao(Earthquake Research Institute, The University of Tokyo), Sergey Kravtsov(University of Wisconsin Milwaukee), Chairperson:Takafumi Niihara(The University Museum, The University of Tokyo), Chairperson:Tatsu Kuwatani(Japan Agency for Marine-Earth Science and Technology)

Sat. May 20, 2017 9:00 AM - 10:30 AM A01 (Tokyo Bay Makuhari Hall)

It is important to extract essential processes and structures from observed datasets in order to understand and predict the dynamic behavior of the earth and planetary systems. Recently, powerful data-driven methodologies have been proposed to extract, model and predict useful information contained in high-dimensional datasets that are ubiquitous in earth and space. This session aims to provide an opportunity to highlight recent advances in such data-driven techniques across disciplines and to have a productive discussion for interdisciplinary collaborations.

[MGI29-01] Application of Cluster Analysis to GNSS Data in the Angular Velocity Space: Identification of Crustal Blocks and Evaluation of Plate Interaction

*Atsushi Takahashi[1], Tomohisa Okazaki[1], Koh Takeuchi[2], Tomoharu Iwata[2], Yukitoshi Fukahata[3] (1.Graduate School of Science, Kyoto University, 2.NTT Communication Science Laboratories, 3.Disaster Prevention Research Institute, Kyoto University)

9:00 AM - 9:15 AM

[MGI29-02] A self-organizing map exploratory analysis of the flow duration curve in the United States

*Geoffrey Fouad[1], André Skupin[2] (1.Monmouth Univ, 2.San Diego State Univ)

9:15 AM - 9:30 AM

[MGI29-03] Uncertainty quantification for groundwater management in the Danish buried valley systems by means of regression tree-based surrogate models

*Jihoon Park[1], Céline Scheidt[1], Jef Caers[2] (1.Department of Energy Resources Engineering, Stanford University, 2.Department of Geological Sciences, Stanford University)

9:30 AM - 9:45 AM

[MGI29-04] Multivariate analysis of visible to near-infrared reflectance spectra of meteorites and asteroids

*Peng Hong[1], Yuria Watabiki[2], Takafumi Niihara[3], Hideaki Miyamoto[1,2,3,4], Yuki Saito[5,6], Kenji Fukumizu[5,6] (1.Department of Systems Innovation, Graduate School of Engineering, The University of Tokyo , 2.Department of Complexity Science and Engineering, Graduate School of Frontier Sciences, The University of Tokyo, 3.The University Museum, The University of Tokyo, 4.Department of Earth and Planetary Science, Graduate School of Science, The University of Tokyo, 5.The Graduate University of Advanced Studies, 6.The Institute of Statistical Mathematics)

9:45 AM - 10:00 AM

[MGI29-05] Relationship between reflectance spectra of meteorites and asteroids visualized by the correlation distance and t-SNE

*Hideaki Miyamoto[1], Takafumi Niihara[1], Peng Hong[1], Hideitsu Hino[2] （1.University of Tokyo, 2.University Tsukuba ）

10:00 AM - 10:15 AM

[MGI29-06] Systematic attribution of observed circulation trends to external forcing and internal variability

*Christian L. E. Franzke[1] （1.University of Hamburg）

10:15 AM - 10:30 AM

Application of Cluster Analysis to GNSS Data in the Angular Velocity Space: Identification of Crustal Blocks and Evaluation of Plate Interaction

*Atsushi Takahashi[1], Tomohisa Okazaki[1], Koh Takeuchi[2], Tomoharu Iwata[2], Yukitoshi Fukahata[3]

1. Graduate School of Science, Kyoto University, 2. NTT Communication Science Laboratories, 3. Disaster Prevention Research Institute, Kyoto University

The motion of a rigid plate on a sphere is expressed as rotation around an axis that passes through the center of the earth. Recently, statistical approaches were proposed to identify boundaries of crustal blocks from observed GNSS data. Simpson et al. (2012) showed that through a cluster analysis of GNSS data, block boundaries can be distinguished objectively. Savage and Simpson (2013) extended the study by adding an iterative algorithm to take account of the effect of spherical geometry. These studies successfully showed a way to identify block structures in the West Coast of the U.S.
However, it is still difficult to apply the method to global GNSS data in order to identify plate blocks considering the effect of spherical geometry. So, we developed a more intuitive method to tackle this problem.
The relationship among the geographical location of the GNSS station, observed velocity at the GNSS station, and candidates of the Euler pole can be expressed as a vector equation: the cross product of an angular velocity vector and a position vector of a GNSS station is an observed velocity vector.
From this relationship, candidates of the Euler pole can be expressed as a straight line in the angular velocity space. We can expect that each line that correspond to each GNSS data in the same rigid crustal block crosses at a point in the angular velocity space.
To spot a crossing point, we made a matrix whose components correspond to the distance between lines. In order to find a structure in the matrix, we analyzed this matrix using a clustering algorithm called a Bayesian Community Detection model. The method provides a block matrix structure within it for a given threshold. By this analysis, we can spot the candidates of Euler poles as a crossing point based on the distances of lines in the angular velocity space. Each identified crossing point would represent a cluster, namely a crustal block.
However, an actual crustal block has internal deformation in it. So, we considered how such deformation affect the deviation of the lines from their original crossing point. We first analyzed the same data set of Simpson et al. (2012) in the San Francisco Bay Area, West Coast of U.S. for comparison.
The obtained result had four major crossing points, which was almost the same as Simpson et al. (2012). However, if we gave a smaller threshold, we obtained 16 minor clusters which reflect internal deformation of crustal blocks. The minor crossing points almost aligns on a straight line that connects the major crossing points.
These minor crossing points can be attributed to the effect of coupling on the faults that bound major blocks in this area. If there is some coupling on such faults, crustal movement must systematically deviate from the rigid block rotation due to internal elastic deformation caused by fault coupling. In other words, we can extract information about fault coupling from this clustering analysis.

A self-organizing map exploratory analysis of the flow duration curve in the United States

*Geoffrey Fouad[1], André Skupin[2]

1. Monmouth Univ, 2. San Diego State Univ

The flow duration curve (FDC) describes the full range of streamflow magnitude observed at a site, and is strongly influenced by upstream conditions of the basin. Upstream conditions are quantified using basin characteristics, such as mean elevation and annual precipitation. A large variety of data now exists to characterize basins in the United States (US). However, a greater understanding of how this data relates to the FDC is critical, considering basin characteristics are typically the basis for predicting the FDC of ungauged basins. The present study performs an exploratory analysis of the FDC and characteristics of 918 basins in the US using a neural network technique called the self-organizing map (SOM). The SOM is applied for its ability to cluster and visualize fine-scale variation in large datasets. Both of these exploratory frameworks (i.e. clustering and visualization) are used to compare individual flows of the FDC to basin characteristics. Clusters based on common basin characteristics poorly agree with those of the FDC (36% agreement), which is less than prior work in smaller study areas, such as Italy. This is an important point because clusters based on basin characteristics are used to deploy models for predicting the FDC. Basin characteristics primarily cluster basins into geographic regions, whereas the FDC generates clusters of basins distributed throughout the US. Geographic proximity therefore may not be an indicator of similarity in the FDC between basins. Variation of the FDC is also unrelated to some common basin characteristics, such as topographic variables, as indicated through SOM data visualizations. This may partially explain the disagreement between the two sets of clusters. The disagreement may also be because basin characteristics are only associated with certain parts of the FDC, but not the overall FDC. For instance, aridity, an index of precipitation lost to evapotranspiration, suppresses high flows possibly due to lower antecedent moisture conditions that moderate storm flows. High flows are also related to spring snowmelt represented using the percent of precipitation delivered as snow. Another association to a part of the FDC is that average to low flows vary with groundwater contributions (i.e. baseflow). Basin characteristics describing surface runoff are more related to high flows, whereas subsurface drainage has more influence on average to low flows. The processes that generate different flows should be accounted for in the clusters used to predict the FDC, and future research should evaluate if the tradition of using a single set of basin characteristics to cluster basins for predicting the FDC should be revised to select different basin characteristics depending on the flow targeted for prediction.

Keywords: basin hydrology, flow duration curve, self-organizing map, cluster analysis, data visualization, United States

Uncertainty quantification for groundwater management in the Danish buried valley systems by means of regression tree-based surrogate models

*Jihoon Park[1], Céline Scheidt[1], Jef Caers[2]

1. Department of Energy Resources Engineering, Stanford University, 2. Department of Geological Sciences, Stanford University

Uncertainty quantification is a key component for decision making in groundwater management. Such applications involve the building of large complex spatial models, the application of computationally intensive forward modeling codes and the integration of heterogeneous sources of uncertainty.
An integral step for uncertainty quantification is to condition models to a variety of data. In the Danish groundwater management this consists of head, streamflow, recharge, well and geophysical (SkyTEM) data. Uncertainty quantification requires model calibration. This is a challenging problem when dealing with complex systems (such as the Danish buried valley system) and a wealth of data. Another difficulty is computational cost, since a proper model calibration should account for all data, all model variables and geological heterogeneity requires running many forward flow models.
In this research, a workflow is proposed to find posterior multivariate distribution of model parameters and predictions. First, dimensionality reduction with mixed principle component analysis (PCA) is performed to incorporate different types of available data. A regression model is built for uncertain model parameters and misfit between simulated and observed data. As a regression model, we use a boosted regression tree because it offers high quality predictive model in nonlinear problems. Another advantage of tree-based approach is that we can obtain predictor importance, which can be directly used in sensitivity analysis.
Models that match the data are found by Approximate Bayesian Computations (ABC), where the likelihood is simply an indicator function of data mismatch. ABC requires exhaustive Monte Carlo sampling and running forward models. By using the regression model as surrogate forward model, we can obtain models conditioned to the data without intensive full forward runs. Regression models can also be constructed for predictions, such as the effect of establishing new wells for extraction.
We illustrate our method using a real field problem of decision making in the Danish groundwater system. Decisions include where to relocate drinking wells while minimizing the change of water produced and effects on farms and industrial areas. Well head and stream data are observed from monitoring wells. The proposed workflow is used to understand the effect of each parameter and to obtain the posterior distribution of 20 forecasts with newly acquired data.

Keywords: uncertainty quantification, regression analysis, model calibration, groundwater management

Multivariate analysis of visible to near-infrared reflectance spectra of meteorites and asteroids

*Peng Hong[1], Yuria Watabiki[2], Takafumi Niihara[3], Hideaki Miyamoto[1,2,3,4], Yuki Saito[5,6], Kenji Fukumizu[5,6]

1. Department of Systems Innovation, Graduate School of Engineering, The University of Tokyo , 2. Department of Complexity Science and Engineering, Graduate School of Frontier Sciences, The University of Tokyo, 3. The University Museum, The University of Tokyo, 4. Department of Earth and Planetary Science, Graduate School of Science, The University of Tokyo, 5. The Graduate University of Advanced Studies, 6. The Institute of Statistical Mathematics

Asteroids and meteorites have been considered as remnants of the early evolution of the solar system and understanding their formation history provide constraints on the physical and chemical conditions of the solar nebula and its subsequent evolution [1]. In order to better understand the compositions of asteroids, mineralogical relationship between asteroids and meteorites have been studied based on reflectance spectra obtained by ground- and space-based telescope [2]. Their relationship, however, remains poorly constrained except for a S-type asteroid Itokawa and LL chondrites [3]. Although spectral similarities have been suggested between V-type asteroids and HED meteorites and between carbonaceous chondrites and C- and/or D-type asteroids, detailed relationship is not well constrained. The major obstacle to compare asteroids and meteorites is that the classification scheme between asteroids and meteorites are fundamentally different. Asteroids are classified mainly based on the shape of their reflectance spectra and orbital parameters [4], while meteorites are classified by detailed petrology and mineralogy [5]. Based on principal component analysis, Britt et al. (1992) [6] compare reflectance spectra of asteroids with those of meteorites. They find that most of principal components of meteorite spectra are offset from those of the bulk of the asteroid population. However they used only eight color spectra and the spectra are limited within visible wavelength from 0.35 to 1.0 μm. Since characteristic absorptions are observed in the near-infrared range, including pyroxene (2 μm) and hydrated silicates (3 μm), using reflectance spectra with a wider wavelength range could result in a better spectral matching between asteroids and meteorites. In this study we developed a database of reflectance spectra for asteroids and meteorites with wavelengths ranging from 0.4 to 4 μm and perform multivariate analysis.

We obtained reflectance spectra for meteorites and asteroids from RELAB [7] and the database of Planetary Spectroscopy at MIT [8], respectively. Asteroid spectra for 3 μm band are obtained from previous studies [e.g., 9]. All the spectra were sampled with cubic spline fits at a wavelength interval of 0.05 μm. Meteorite spectra are chosen based on the following criteria: (1) particulate sample, (2) phase angle is 30°, (3) sample is from valid/known meteorite, (4) not heated/laser-irradiated, inclusion or impact melt sample, (5) not moon sample or lunar meteorite. The developed database includes 534 meteorite spectra and 369 asteroid spectra. We performed principal component analysis on the database and measure how well each meteorite group and asteroid group is separated on the principal component space. Our preliminary analyses show that (1) using spectra from 0.4 to 2.5 μm, accuracy of separation among ordinary chondrites, carbonaceous chondrites, HED meteorites is significantly improved compared with the case using spectra from 0.4 to 1.0 μm, and (2) the accuracy of separation is not significantly improved when using meteorite spectra from 0.4 to 4 μm compared with the case using spectra from 0.4 to 2.5 μm.

References: [1] DeMeo F.E. & Carry B. (2014) Nature, 505, 629. [2] Pieters C.M. & McFadden L.A. (1994) Annu. Rev. Earth Planet. Sci., 22, 457. [3] Nakamura, T. et al. (2011) Science, 333, 1113. [4] DeMeo F. E. et al. (2009) Icarus, 202, 160. [5] Weisberg M.K. et al. (2006) Meteorites and the Early Solar System II, 19.

[6] Britt D. T. et al. (1992) Icarus, 99, 153. [7] Pieters C.M. & Hiroi T. (2004) LPSC XXXV, 1720. [8] http://smass.mit.edu/home.html [9] Rivkin A.S. et al. (2000) Icarus, 145, 351.

Keywords: asteroids, meteorites, reflectance spectra

Relationship between reflectance spectra of meteorites and asteroids visualized by the correlation distance and t-SNE

*Hideaki Miyamoto[1], Takafumi Niihara[1], Peng Hong[1], Hideitsu Hino[2]

1. University of Tokyo, 2. University Tsukuba

Classification of asteroids based on orbits, sizes, and reflectance spectra have been performed for decades to understand the nature of small bodies. As for asteroids, asteroid taxonomic systems based on color, albedo, and spectral shape have been developed and modified/expanded to grasp their variations. While each spectral class is believed to represent a specific composition of asteroids, their correlation is still not fully understood because of the nature of reflectance spectrum of rock-forming minerals; The relationship between reflectance spectra of an asteroid and a meteorite is essentially difficult to unravel without detailed analyses of the shapes of spectra. Even so, several previous attempts exist to statistically solve this issue. For example, Britt et al (1992) successfully produced a map of statistically defined spectral similarities and found that principal component analysis is successful at characterizing the primary spectral variance in the asteroid and meteorite populations. In other words, statistical classifications of spectral types without detailed interpretation of spectral shapes can be useful to overview the variation and relationships within a spectral data set, even though there are known difficulties of comminution, melting, mixing, and space weathering. In this work, we expand the above idea by applying to a wider and denser datasets of reflectance spectra for both meteorites and asteroids. We use published databases of RELAB's laboratory measurements of meteorites and Planetary Spectroscopy at MIT's asteroid spectra, which are resampled by cubic spline fits in the wavelengths ranging from 0.45 to 2.45 μm with the wavelength resolution of 0.05. We statistically analyze the distance of spectra by means of such as Partial Autocorrelation, Dynamic Time Warping, Pearson Correlation, and Euclidean distance. Results are visualized by using 6 kinds of schemes including t-SNE (t-Stochastic Neighbor Embedding). We find that correlations of both meteorites and asteroids are generally shown by this simple scheme. Preliminary results indicate that (1) V-type asteroids generally match HED meteorites, (2) S-type asteroids locate near ordinary chondrites but they do not entirely match each other, which may reflect the effect of space weathering (3) C-type asteroids match carbonaceous chondrites and they are separated into a few sub clusters.

Keywords: asteroid, meteorite, reflectance spectra

Systematic attribution of observed circulation trends to external forcing and internal variability

*Christian L. E. Franzke[1]

1. University of Hamburg

The atmospheric circulation can be efficiently described by just a few regime states or teleconnection patterns. For the systematic identification of these regime states a novel space-time clustering method has been developed (FEM-BV-VARX). This method identifies persistent regime states which are important for predictions. In my presentation I will discuss the use of this method for the attribution of circulation trends and extreme events.

Keywords: Regime States, Circulation Trends, Extreme Events, Clustering Method

